CORTEX 74 (2016) 79–95

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Is less really more: Does a prefrontal efficiency genotype actually confer better performance when working memory becomes difficult?

Jessica L. Ihne [a], *Natalie M. Gallagher* [a], *Marie Sullivan* [a], *Joseph H. Callicott* [b] *and Adam E. Green* [a,*]

[a] *Department of Psychology, Georgetown University, United States*
[b] *Clinical Brain Disorders Branch, National Institute of Mental Health, United States*

ARTICLE INFO

Article history:
Received 22 May 2015
Reviewed 10 August 2015
Revised 24 September 2015
Accepted 29 October 2015
Action editor Ahmad Hariri
Published online 14 November 2015

Keywords:
COMT
Working memory
Neural efficiency
Dopamine
DLPFC

ABSTRACT

Perhaps the most widely studied effect to emerge from the combination of neuroimaging and human genetics is the association of the COMT-Val$^{108/158}$Met polymorphism with prefrontal activity during working memory. COMT-Val is a putative risk factor in schizophrenia, which is characterized by disordered prefrontal function. Work in healthy populations has sought to characterize mechanisms by which the valine (Val) allele may lead to disadvantaged prefrontal cognition. Lower activity in methionine (Met) carriers has been interpreted as advantageous neural efficiency. Notably, however, studies reporting COMT effects on neural efficiency have generally not reported working memory performance effects. Those studies have employed relatively low/easy working memory loads. Higher loads are known to elicit individual differences in working memory performance that are not visible at lower loads. If COMT-Met confers greater neural efficiency when working memory is easy, a reasonable prediction is that Met carriers will be better able to cope with increasing demand for neural resources when working memory becomes difficult. To our knowledge, this prediction has thus far gone untested. Here, we tested performance on three working memory tasks. Performance on each task was measured at multiple levels of load/difficulty, including loads more demanding than those used in prior studies. We found no genotype-by-load interactions or main effects of COMT genotype on accuracy or reaction time. Indeed, even testing for performance differences at each load of each task failed to find a single significant effect of COMT genotype. Thus, even if COMT genotype has the effects on prefrontal efficiency that prior work has suggested, such effects may not directly impact high-load working memory ability. The present findings accord with previous evidence that behavioral effects of COMT are small or nonexistent and, more broadly, with a growing consensus that substantial effects on phenotype will not emerge from candidate gene studies.

© 2015 Elsevier Ltd. All rights reserved.

* *Corresponding author*. Department of Psychology, Georgetown University, 302C White-Gravenor Hall, 3700 O Street, NW, Box 571001, Washington, DC 20057, United States.
E-mail address: aeg58@Georgetown.edu (A.E. Green).
http://dx.doi.org/10.1016/j.cortex.2015.10.025
0010-9452/© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

A challenge that has characterized the early years of cognitive neurogenetic research has been finding mechanistic links from the way a genetic variant affects the brain to true effects on behavior (Green et al., 2008). Perhaps the most widely studied genetic variant in this burgeoning literature has been the Val$^{108/158}$Met polymorphism of the dopamine system-related catechol-O-methyltransferase (COMT) gene. The association of the COMT-Met allele with relatively reduced dorsolateral prefrontal cortex (DLPFC) activation during working memory is among the best-replicated cognitive neurogenetic findings to emerge thus far (see Mier, Kirsch, & Meyer-Lindenberg, 2010, for review). The COMT Val allele has been identified as a possible schizophrenia risk factor (Egan et al., 2001; Glatt, Faraone, & Tsuang, 2003; but see Munafo, Bowes, Clark, & Flint, 2005). Work in healthy populations has sought to inform understanding of schizophrenia diathesis by characterizing general mechanisms by which COMT genotypes affect prefrontal dopaminergic neurocognition (Winterer & Weinberger, 2004). Reduced DLPFC activity in Met carriers has been interpreted as greater neural efficiency relative to Val allele carriers (Egan et al., 2001). This account is appealing based on molecular biological evidence that the COMT-Met allele is associated with less inactivation of dopamine in PFC, presumably enabling more efficient dopaminergic signaling (Chen et al., 2004; Lachman et al., 1996). However, a key aspect of the COMT efficiency account has gone largely unexplored. Specifically, if Met allele carriers process working memory demands more efficiently in DLPFC, these individuals should be better able to cope with increasing demands on DLPFC resources. If this interpretation is correct, it should be evident in better performance for Met carriers than Val carriers when working memory tasks become highly demanding. Higher working loads in working memory and related tasks have been shown to elicit individual differences in performance that are not evident at lower loads (Bielinski & Davison, 1998; Drew & Vogel, 2008; Grabner et al., 2007; Kane & Engle, 2003; Linden et al., 2003; Todd & Marois, 2004). However, to our knowledge, no study has yet explored the effects of COMT genotype (or, indeed, any genetic variant) on performance at highly demanding working memory loads. Addressing this unresolved issue was the focus of the present study.

1.1. *COMT and prefrontal dopaminergic signaling*

The COMT gene codes for the COMT enzyme, which inactivates catecholamines, including dopamine, in the synaptic cleft (Axelrod & Tomchick, 1958). In prefrontal cortex, COMT is the primary regulator of synaptic dopamine levels, since PFC exhibits very low density of dopamine transporters (DAT), which are responsible for clearance of dopamine from the cleft in other brain regions (Käenmäki et al., 2010; Sesack, Hawrylak, Matus, Guido, & Levey, 1998). Within the COMT gene, the single nucleotide polymorphism (SNP) at amino acid 108/158 (rs4680) has a substantial effect on COMT enzyme activity, and thus the amount of available synaptic dopamine in PFC. The COMT Val$^{108/158}$Met SNP produces a valine-to-methionine substitution (Lotta et al., 1995). The valine (Val) allele is more heat stable and has three to four times higher enzyme activity than the methionine (Met) allele (Lachman et al., 1996). This results in the inactivation of more dopamine and ultimately lower levels of synaptic dopamine associated with the Val allele. Necessarily, the opposite effects (relatively lower enzymatic activity and greater dopamine availability) are associated with the Met allele (Chen et al., 2004).

Previous research has demonstrated the relationship between prefrontal dopamine and higher-level cognitive abilities (D'Esposito et al., 1995; Williams & Goldman-Rakic, 1995). Working memory, a pillar of higher cognition (Ackerman, Beier, & Boyle, 2005), has been linked to neurotransmission via the dopamine D_1 receptor and the balance between D_1 and D_2 receptor binding in PFC (Goldman-Rakic, Muly, 3rd, & Williams, 2000; Winterer et al., 2006). Pharmacological manipulation via the COMT inhibitor, tolcapone (Dingemanse et al., 1995) has been shown to result in higher levels of prefrontal dopamine. Tolcapone has been linked to improved working memory (Farrell, Tunbridge, Braeutigam, & Harrison, 2012; Giakoumaki, Roussos, & Bitsios, 2008) and other executive functioning (Apud et al., 2007). There is some evidence to indicate that tolcapone-related cognitive benefits are specific to individuals with COMT Val/Val genotype, who would normally have lower synaptic dopamine in the PFC than Met allele carriers. However, these findings have been somewhat inconsistent. Apud et al. (2007) found no COMT genotype X tolcapone treatment interaction effects on performance of the working memory tasks they tested. Giacoumaki et al. (2008) found that a genotype X tolcapone interaction was not significant after controlling for IQ. Farrell et al. (2012) found significantly better performance for Met/Met homozygotes than Val/Val homozygotes at 0-back, 1-back, 2-back, and 3-back conditions of the N-back task in the placebo group, which bolstered the genotype X tolcapone effect, but is contrary to the majority of prior evidence that performance does not differ at low working memory loads (Apud et al., 2007; Congdon, Constable, Lesch, & Canli, 2009; Dennis et al., 2010; de Frias et al., 2010). In addition, the studies that reported genotype X tolcapone interaction effects on working memory performance (Farrell et al., 2012; Giacoumaki et al., 2008) did not account for possible effects of ethnic stratification and were based on relatively small samples (genotype groups between 11 and 18 participants).

1.2. *Evidence and questions for the COMT efficiency account*

Multiple brain-imaging studies have found differences in prefrontal activity associated with variation in COMT genotype (see Mier et al., 2010, for review), consistent with the role of the COMT enzyme as a primary dopamine regulator in mammalian PFC. Functional magnetic resonance imaging (fMRI) has been used to measure differences in task-related brain activation between COMT genotype groups. An initial study by Egan et al. (2001) found that, during low-load working memory, individuals with the Val/Val genotype showed greater prefrontal activation than those with Val/Met, who

showed greater activation than Met/Met individuals. The association between COMT-Val and relatively greater levels of working memory-related activation has since been well-replicated, and demonstrated via meta-analysis (Mier et al., 2010). Because Val carriers appear to use more DLPFC activation to achieve the same working memory performance that Met carriers achieve with less activation, the fMRI data have been interpreted as evidence of inefficient PFC functioning associated with the Val allele (Caldu, et al., 2007; Egan et al., 2001; Mattay et al., 2003; Tan, Chen, Goldberg, et al., 2007).

Based on this COMT efficiency account, it is reasonable to predict that the greater neural efficiency of Met carriers will lead to a cognitive performance advantage when greater demands are placed on PFC working memory resources. However, the question of whether COMT genotypes associated with differences in fMRI activation actually confer appreciable differences in working memory ability has been insufficiently addressed. Imaging studies of COMT genotype have used relatively low working memory loads, and have generally found equivalent levels of performance across genotypes (Congdon et al., 2009; Dennis et al., 2010; de Frias et al., 2010). In addition, these studies have tended to include only highly accurate healthy volunteers because BOLD fMRI requires a sufficiently large number of correct trials to power data analysis. Thus, while the evidence is rather clear for the link between COMT and neural activation, associations with actual cognitive performance/ability have been less clear.

Behavioral investigations of the influence of COMT on cognitive abilities have generally used the same tasks implemented in fMRI COMT studies and, critically, have also applied relatively low levels of working memory load. Some research has shown the Met allele to be associated with better performance on working memory (Diaz-Asper et al., 2008), the Wisconsin Card Sorting task (Egan et al., 2001), and IQ (Green, Kraemer, DeYoung, Fossella, & Gray, 2013). Though other investigations have found COMT genotype associations with working memory performance to be inconsistent (Barnett, Scoriels, & Munafò, 2008; Bruder et al., 2005; Raz, Dahle, Rodrigue, Kennedy, & Land, 2011), or have found no significant differences between genotypes on cognitive performance measures (Wardle, de Wit, Penton-Voak, Lewis, & Munafò, 2013), even in the presence of lower prefrontal activity in healthy Met carriers (Apud et al., 2007; Dennis et al., 2010). Still other studies have shown the Met allele to be associated with less activation in some prefrontal brain regions, yet greater activation in other areas (Congdon et al., 2009; Green et al., 2013; Winterer et al., 2006; de Frias et al., 2010). This inconsistency is set within a broader context of ambiguity about the interpretation of greater versus lesser neural activity with respect to cognitive efficacy and "efficiency" (Poldrack, 2014). Notably, better executive cognition has sometimes been associated with increased, rather than decreased, prefrontal activity (Gray, Chabris, & Braver, 2003; Osaka et al., 2004). In view of the uncertainties surrounding the link between COMT genotype and cognitive function, it is important to consider the possibility that COMT genotype-associated differences in neural activity at low working memory loads reflect differences in neural recruitment that do not bear directly on differences in cognitive ability.

1.3. *Approaches to measuring working memory demands*

Among studies that have investigated the influence of COMT genotype on working memory, the N-back task paradigm has been frequently used. N-back tasks present a serial sequence of items (e.g., numbers, letters, words, images). For each item presented as the sequence proceeds, the participant is required to recall and make a judgment about the item that was presented a specified number of items (N) previously. N-back tasks involve multiple cognitive functions, such as storing information, constantly updating this storage, and inhibiting distractions (Owen, McMillan, Laird, & Bullmore, 2005). There are multiple versions of the N-back task, which place varying demands on storage, inhibition, and manipulation and involve different modalities of information to be held in working memory (e.g., item identities, locations of items in space). N-back tasks are suited to testing across multiple working memory loads, e.g., 0-back, 1-back, 2-back, etc. Accuracy is consistently very high for 0-back and 1-back. Performance shows relatively small decreases as load increases to 2-back, and performance at 3-back remains very high relative to chance (Callicott et al., 1999; Jonides et al., 1997; Nystrom et al., 2000). Prior evidence indicates that increasing working memory load above three items yields individual differences in performance that are not otherwise visible (Linden et al., 2003; Todd & Marois, 2004). Thus, it is possible that neuroimaging studies that have tested working memory performance differences between COMT genotypes may have used tasks that were not difficult enough to elicit sufficient behavioral differences.

Another working memory paradigm that has been used to investigate PFC activation, though not as frequently as the N-back, is the classic Sternberg task (Sternberg, 1966). In this paradigm, a memory set of items (often alpha-numeric characters) are presented, followed by a delay and then a single probe item. When the probe item appears, the participant responds to indicate whether the probe item was in the memory set. The Sternberg task involves substantial demand on the storage capacity of working memory for multiple items (especially for large memory sets), but relatively less demand on inhibition because the delay period between probe and target does not involve distractors. Neuroimaging studies of the Sternberg task have found increasing PFC activation driven by increasing working memory load (i.e., a greater number of items in the memory set) (Altamura et al., 2007; Veltman, Rombouts, & Dolan, 2003). In Sternberg variations, the highest loads tested generally have not exceeded 6 or occasionally 8 stimuli in a memory set. As with the N-back task, accuracy has been shown to be quite high even at the most difficult loads used, and consistently well above a priori rates of chance performance.

Though there is not a 'gold standard' single task assay of working memory, both Sternberg and n-back tasks are well-established tasks that have been shown to be useful assessments of working memory capacity (Barch 2012; Wilhelm 2013). These tasks have been widely used in neuroimaging studies that have demonstrated the role of the prefrontal cortex in working memory, supporting a connection to the

important role of COMT in PFC. Nonetheless, because working memory is a multifaceted construct, it is informative to test working memory tasks with varying neuropsychological characteristics simultaneously and to apply factor-analytic approaches that can enable insights about working memory as a construct underlying performance across tasks.

In order to determine whether any COMT genotype-related effects apply to general working memory ability, rather than to performance on a specific task with unique demand characteristics, the present investigation utilized multiple working memory tasks within the same cohort. Thus, the present investigation was also positioned to measure any genotype-related performance differences that might be associated with differing cognitive demands of different working memory tasks. This multi-task approach was motivated by the fact that many of the empirical observations that support the COMT efficiency account have been based on a single version of the N-back paradigm (Bertolino et al., 2006; Egan et al., 2001; Mattay et al., 2003; Tan, Chen, Sust, et al., 2007). This version of N-back involves several unique characteristics (e.g., simultaneous representation of item identity and spatial location). Consideration of the informational content of working memory tasks is particularly important in view of evidence that such content (particularly spatial *vs* nonspatial information) impacts the relationship of working memory performance to changes in measures of neural efficiency due to practice (Sayala, Sala, & Courtney, 2006). Recent evidence of domain specificity in working memory storage, even at high working memory loads (Fougnie, Zughni, Godwin, & Marois, 2015), further underscores the utility of considering working memory across multiple tasks with varying characteristics, rather than a single working memory task.

Testing the appealing but empirically unexplored prediction that increasing working memory demand will increase differences in performance between COMT genotype groups is critical to understanding whether and how the COMT-Met allele confers an actual cognitive advantage. For example, on an N-back task, if there are no performance differences between genotypes on 2-back or 3-back, will there be slight differences on 4-back, and greater differences on 5-back? In the present investigation, we tested participants on three classical working memory tasks that were modified to include multiple levels of load, including loads that were higher (more difficult) than those used in previous studies of COMT genotype. We aimed to test whether any differences in performance between COMT genotypes would emerge as load increased.

2. Material and method

2.1. *Participants*

Participants were 111 caucasian healthy volunteers, between the ages of 18 and 35. We restricted our analyses to caucasian participants in order to avoid ethnic stratification, consistent with prior COMT research (Egan et al., 2001; Raz et al., 2011), and because of the different allele frequencies of the COMT polymorphism across ethnicities (Palmatier, Kang, & Kidd, 1999). Power calculation to determine sample requirements used an effect size estimated from a previous study (Raz et al., 2011). An ideal basis for the effect size estimate would be a study reporting a COMT genotype by working memory load interaction in healthy young participants. To our knowledge, no such study has yet been conducted, which contributes to the motivation for the present study. Raz et al. (2011) report a three-way interaction effect between working memory load, genotype for COMT-Val$^{108/158}$Met, and genotype for an insertion/deletion polymorphism of the angiotensin converting enzyme (ACE) gene, in a sample that included adults as old as 81 years, and only investigating 1-Back, 2-Back, and 3-Back loads. The f^2 effect size for that interaction was .11 (Raz et al., 2011). In order to detect a 3 (genotype) X 4 (load) effect of this size in the present study at $\alpha = .05$, a total sample of 36 would yield 95% power. Because this calculation was based on a model that was non-identical to the planned model in the present study, because effects might be smaller in a healthy young cohort than in older adults, and most importantly because we sought sufficient statistical power to identify the absence of an effect at even a trend level, we deemed that exceeding the calculated minimum sample size by an approximately three-fold increase would be appropriate.

Munafo and Flint (2007) conducted an important meta-analysis of the COMT literature, which did not support a difference between COMT genotypes in N-back performance. That meta-analysis included studies that tested relatively low levels of working memory load (primarily 1-back and 2-back, with no study exceeding 3-back), as those were the only studies available in the literature. While our study includes loads as low as 2-back, the focus of our study is on working memory at high loads (4-back and 5-back in the N-back task, and 8- and 10-letter memory sets in the Sternberg task), which has not yet been investigated with respect to COMT. Since the loads we are exploring are new in the context of genetic research, the extant literature does not provide examples upon which to base clear calculation of required sample size based on power. Nonetheless, our sample size afforded 80% power to detect a small effect size of $f^2 = .01$ at $\alpha = .05$ for a genotype-by-load interaction. A negative finding in the present study cannot eliminate conclusively that some effects may have been missed due to sample size. However, our power estimates indicate that if any effect exists it is unlikely to exceed a small effect size. As described below, our results generally did not indicate even trend level significance (even relative to an uncorrected significance threshold of $\alpha = .05$), so it is not likely that unobserved effects were present close to this study's sensitivity to detect them. It is also worth noting that the present study included a sample size that was generally larger than the samples used in the COMT neuroimaging literature that this study is intended to inform.

The three genotype groups in our sample did not differ in average age, $F(2, 108) = .02$; $p = .979$, nor in the distribution of males and females, X^2 (2, $N = 111$) = .64; $p = .727$, and no such age or sex differences were present in the participants who's data met quality control criteria for all three tasks (described below). Participants were recruited through posted advertisements and through the Georgetown Research Volunteer Program. All participants gave written, informed consent in accordance with the Georgetown University Institutional Review Board.

2.2. Working memory tasks

Participants were tested on three different working memory tasks, with the task order randomized and counterbalanced across all participants. Tasks were programmed using E-Prime software (Psychology Software Tools, Inc.) and were designed to test participants across a range of difficulty loads, including loads that were more difficult than those commonly used in prior research on COMT. Before each task, participants were given instructions about how to do the task and were trained via a shortened practice version of each task.

2.2.1. Spatial N-back

The original Gevins-type N-back task with added spatial component (Callicott et al., 1998; 1999) was here called the 'Spatial N-back'. Although both spatial and verbal strategies can be used to perform this task, we use the term 'spatial' only to differentiate from the other N-back task we also tested (below). Previous research using this version of the N-back task has been the basis for much of the literature on differences in brain activation between COMT genotypes (Egan et al., 2001; Mattay et al., 2003). In this Spatial N-back, the numbers 1–4 were presented visually (one at a time for 500 msec, every 2 sec) at the points of a diamond (Fig. 1). Each of the four numbers was always presented in the same location on the diamond: 1 on the top, 2 on the left, 3 on the right, and 4 on the bottom. Response keys for the task were arranged in a diamond configuration, matching the spatial locations of the stimuli. Each time a number was presented, participants pressed the key on the keyboard corresponding to the number seen *N trials previously*. Participants were tested at loads of N = 2, 3, 4, and 5 (i.e., 2-back, 3-back, 4-back, and 5-back), with the level of difficulty increasing as N increased. For each of the four loads, there were 45 total trials presented in a pseudorandom continuous sequence.

2.2.2. Sequential N-back

The Cohen-like N-back task (Cohen et al., 1997; Jonides et al., 1997) was here called the 'Sequential N-back'. In our version of this task, upper-case letters (consonants only) were pseudorandomly presented on the screen, one at a time every 2 sec for a duration of 800 msec (Fig. 2). Participants were instructed to press one key if the letter on the screen was the same as the letter displayed N screens back, and another key if it was not, essentially answering yes or no for whether the current letter matched the letter N-back in the sequence. We administered 2-back, 3-back, 4-back, and 5-back loads. There were 21 key responses (yes or no decisions) required at each load, resulting in a slightly increased number of letters presented for each increasing load: 23 for 2-back (since there are no responses made for the first 2 stimuli), 24 for 3-back, 25 for 4-back, and 26 for 5-back.

2.2.2.1. Sternberg Task. A version of the Sternberg working memory task (Sternberg, 1966) was used, in which memory sets of simultaneously presented upper-case letters (consonants only) were displayed onscreen, followed by a fixation dot, and then a single probe letter (Fig. 3). Participants were instructed to press one key if the probe letter had appeared in the memory set, and another key if it had not. Each memory set was displayed for 4 sec, and each fixation dot (delay) was displayed for 6 sec. The size of the memory sets varied, comprising 4, 6, 8, or 10 letters, with working memory load increasing as the number of letters in the memory set increased. There were 12 trials for each load, all pseudorandomized.

2.3. Genotyping

After each subject completed behavioral testing, we collected a saliva sample using either an Oragene (OG-500) DNA Collection Kit (DNA Genotek Inc.) or a Norgen (RU35700) Saliva DNA Collection, Preservation, and Isolation Kit (Norgen Biotek Corp.). DNA was extracted from saliva, and genotyping was carried out using TaqMan® SNP Genotyping Assays according to the manufacturer's protocol (Applied Biosystems, Inc.). The DNA samples were amplified with the Allelic Discrimination Protocol on an ABI 7900HT system and SDS software using rs4680 primer/probe sets for COMT $Val^{108/158}$Met. Previously analyzed DNA samples with established genotypes representing each possible COMT genotype (Val/Val, Val/Met, Met/Met) were also tested on all reaction plates as standards for quality control. All runs included negative controls without DNA template. Based on COMT gene standards we obtained 100% correct calls and ≥95% quality value. For quality control, genotyping was repeated on about 20% of the samples, resulting in 100% recall.

3. Calculation

For the working memory tasks, performance was measured by accuracy (percent correct) and reaction time, with an average

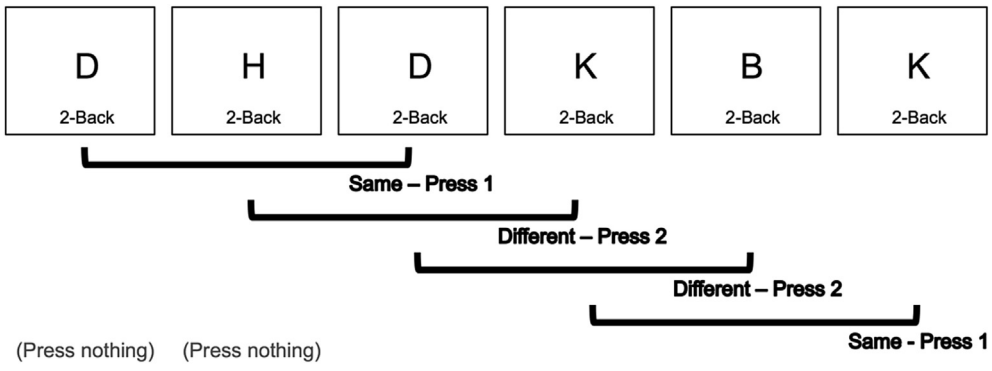

Fig. 1 – Schematic representation of the Spatial N-Back task, showing correct responses for trials at a 2-back load.

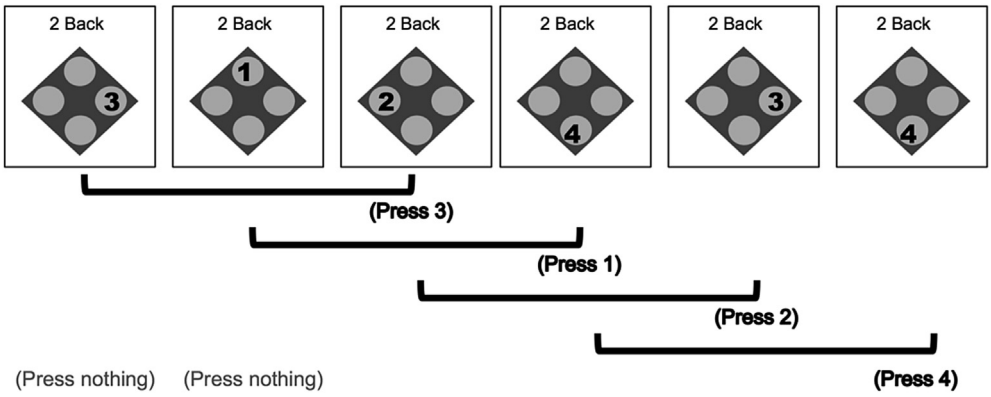

Fig. 2 – Schematic representation of the Sequential N-back task, showing responses for a 2-back load.

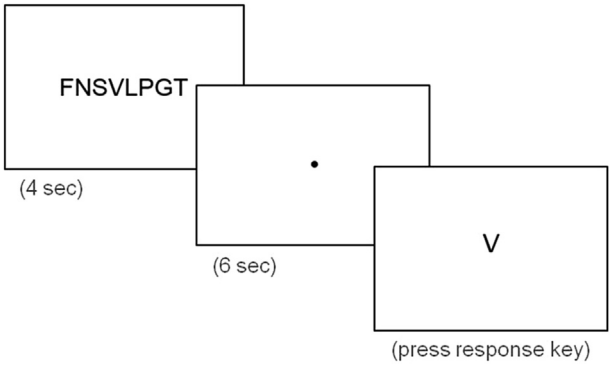

Fig. 3 – Schematic representation of the Sternberg task, showing an 8-letter load.

value across all trials calculated for each load of each task. We grouped participants by COMT genotype (Val/Val, Val/Met, and Met/Met) and tested whether performance differences emerged between genotypes as load increased. To do this, we used a 3-by-4 (genotype-by-load) repeated measures ANOVA for each task to test for a significant interaction effect, as well as for main effects of COMT genotype and working memory load. Because we ran ANOVAs for accuracy and reaction time for each of the three tasks, we corrected for multiple comparisons via Bonferroni correction. We also analyzed the data across all three tasks by converting the accuracy and reaction time measures to standardized z-scores at each load, and then we averaged these values across the tasks for each of the four loads. Additionally, we analyzed the effect of genotype across loads and across tasks with the averaged Z-scored values in a 2-way repeated measures ANOVA to analyze genotype-by-load-by-task effects.

Because working memory is a broad construct, which is not captured by a single gold-standard task, we undertook a factor-analytic approach in order to determine whether performance on the three tasks could be characterized in terms of a working memory factor or factors. Such a cross-task factor would support inferences closer to the construct level of working memory.

Performance on working memory tasks is often reported using the sensitivity index (d′). Of the tasks used in our study, d′ was only an appropriate measure for the Sequential N-back task. d′ is not appropriate for the Spatial N-back because this task does not include the "false alarm" responses that are required for the d′ calculation. Rather than indicating whether a presented number occurred N trials previously in a Yes/No manner, all Spatial N-back responses indicate a number/position that occurred in the stimulus presented N trials previously. Thus, answers can be incorrect, but cannot be classified as false alarms. Performance on the Sternberg task is generally reported as accuracy (% correct) rather than d′ (e.g., Gladwin, den Uyl, Fregni, & Wiers, 2012; Howard et al., 2003; Raghavachari et al., 2001; Zakrzewska & Brzezicka, 2014). This is likely because the ability that the Sternberg task measures (i.e., the ability to hold a large number of items in mind simultaneously despite a delay) is theoretically more important for FALSE trials (i.e., trials on which the correct answer is No) than TRUE trials (i.e., trials on which the correct answer is Yes). On FALSE trials the participant has to remember all the letters in the memory set in order to compare each letter to the probe and determine that each one was not in the memory set, and thus to correctly answer, "No." By contrast, on TRUE trials, it is possible that some correct responses may result from the participant remembering only a portion of the memory set that includes the probe item even though other items in the memory set were not successfully remembered. Thus, d′, which generally treats TRUE trials as "signal" trials and FALSE trials as "noise" trials is not appropriate. Because the Sternberg task required answers on all trials (whether the correct answer is Yes or No), accuracy as % correct accounts for the noise of false "Yes" responding and false "No" responding. Furthermore, the numbers of TRUE and FALSE trials were equivalent in our implementation of the Sternberg task. A primary motivation for using d′ is to avoid overvaluing frequency matching (e.g., guessing "Yes" most of the time when the majority of trials are TRUE). d′ adds far less value when TRUE/signal and FALSE/noise trials are balanced (McNicol, 1972).

Because the previous research to which we sought to compare the present study has frequently shown high cognitive performance across all COMT genotypes at low loads, we specifically examined performance for high performers. For each task, we classified as high performers individuals who demonstrated above 90% accuracy on the

easiest load of the task (i.e., the 2-back for the N-back tasks and the 4-letter load for Sternberg). We then conducted 3-by-4 (genotype-by-load) ANOVAs for accuracy and reaction time using these high-performing groups. For repeated measures ANOVA analyses in which sphericity was violated at $p < .05$ on Mauchly's Test of Sphericity, Greenhouse-Geisser adjustments were used and reported in degrees of freedom, F, and p values.

4. Results

Participants were genotyped for the COMT Val$^{108/158}$Met polymorphism. The numbers of participants per genotype group included in analyses for full samples and for high-performing samples are shown for each task in Table 1. The allelic frequencies for the total sample were .47 for the Val allele and .53 for Met. These frequencies were confirmed to be in Hardy–Weinberg equilibrium ($p = .145$), and also met population norms, based on previous reported frequencies in Caucasian populations that are the same or similar to our observed frequencies: .47 Val and .53 Met (Lee 2014); .48 Val and .52 Met (Palmatier et al., 1999). We analyzed accuracy (percent correct) and reaction time measures for each of the four loads of each working memory task. We used these data to determine whether any differences in performance between COMT genotypes emerged as load became more difficult. ANOVA revealed no genotype-by-load effects on accuracy or reaction time for any of the tasks, and no main effects of COMT genotype on performance.

4.1. *Overall effects across tasks*

To investigate overall main and interaction effects of genotype and working memory demand, performance data were combined across the three tasks. After quality control (detailed for each task below) participants whose data were retained for all three tasks ($N = 99$) were included in this overall analysis. Performance showed a high degree of correlation across the three tasks (Supplementary Tables 1 and 2). Accuracy and RT data were standardized (z-scored) within each load of each task. Averages of these z-scores were then calculated for each load level across tasks. For example, an average was computed for the first load level across tasks by averaging z-scored data from the 2-back condition of the Spatial N-back, the 2-back condition of the Sequential N-back, and the 4-letter condition of the Sternberg task. Using these standardized data, a task-by-load-by-genotype mixed ANOVA revealed no overall main effect of genotype on accuracy, $F(2, 96) = .50$; $p = .607$, or reaction time, $F(2, 96) = .77$; $p = .467$, and no genotype-by-load interaction effects on accuracy, $F(6, 190) = .85$; $p = .531$ (Fig. 4), or reaction time, $F(6, 190) = .74$; $p = .622$ (Fig. 5). As expected due to averaging z-scores, there were no main effects of load on accuracy, $F(3, 94) = .05$; $p = .983$, or reaction time, $F(3, 94) = .13$; $p = .940$. No main effects of task on accuracy, $F(2, 95) = .02$; $p = .978$, or reaction time, $F(2, 95) = .16$; $p = .855$ were observed. There was no genotype-by-task effect on accuracy, $F(4, 192) = .09$; $p = .974$. A genotype-by-task effect on reaction time, $F(4, 192) = 2.52$; $p = .042$, indicated nominal significance, but not corrected significance. There were no load-by-task interaction effects on accuracy, $F(6, 91) = .37$; $p = .898$, or reaction time, $F(6, 91) = .08$; $p = .998$. There also were no three-way interactions (genotype-by-load-by-task) for accuracy $F(12, 184) = 1.11$; $p = .358$, or reaction time, $F(12, 184) = .58$; $p = .857$. Comparisons between tasks were not of primary interest in the present study and such comparisons should be interpreted cautiously because differing a priori rates of chance accuracy across tasks may exaggerate differences, while standardizing data within tasks is likely to minimize differences between tasks.

Because performance at the two highest working memory loads we employed is less well-characterized in the working memory literature than performance at lower-loads, we tested internal reliability on high-load task trials. The split–half reliability (using Spearman–Brown correction) for odd versus even numbered high-load trials across the three tasks was .92, indicating high reliability.

Cross-task factor analysis was conducted using three variables (one variable for each task averaged across loads). This was done four times: accuracy across all four loads, response time across all four loads, accuracy across the two most difficult loads (e.g., 4-back and 5-back), and response time across the two most difficult loads. In each model, a single factor was found with an eigenvalue above 1 (in all models, the eigenvalue for the most explanatory factor > 1.43, %47.65 variance explained). All tasks loaded positively on the most explanatory factor in each model (all factor loadings > .496). This finding confirms our expectation that the three tasks would successfully tap a common cognitive factor, which may be reflective of working memory at the construct level. Regression factor scores for the most explanatory factor in each model were saved out, and used to compare between genotype groups, revealing no differences (see Supplementary Results).

4.2. *Spatial N-Back*

Quality control filtering was applied to remove from all Spatial N-back analyses participants who's performance did not exceed the a priori chance rate of accuracy (25% correct

Table 1 – Numbers of participants by genotype for the full samples and the high-performing samples for each task (numbers of males indicated in parentheses).

	Spatial N-back		Sequential N-back		Sternberg	
COMT	All	High Perf.	All	High Perf.	All	High Perf.
Val/Val	21 (11 M)	4 (4 M)	20 (10 M)	13 (7 M)	20 (10 M)	18 (10 M)
Val/Met	58 (26 M)	12 (8 M)	62 (28 M)	28 (17 M)	60 (25 M)	55 (27 M)
Met/Met	27 (14 M)	6 (3 M)	27 (14 M)	16 (10 M)	25 (13 M)	23 (13 M)

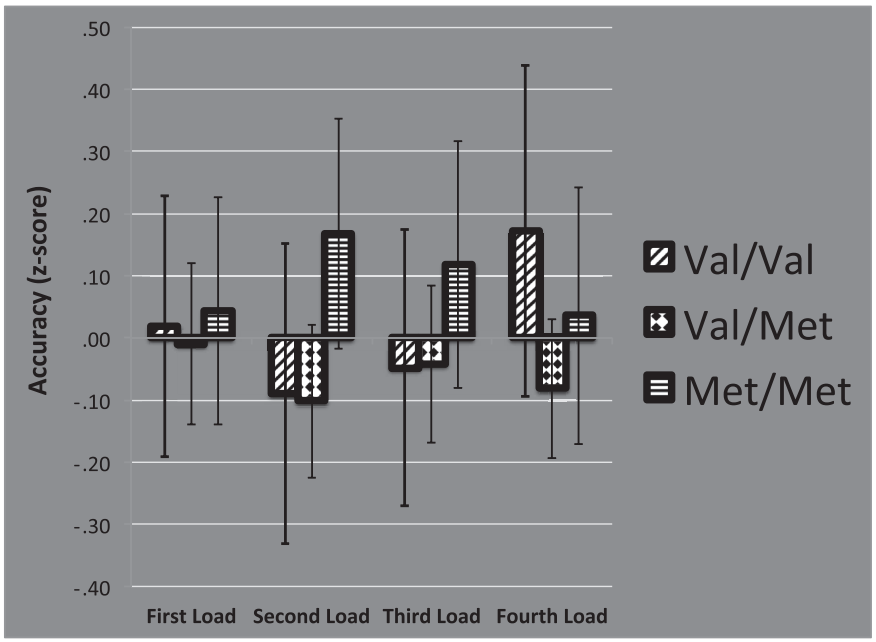

Fig. 4 – Standardized accuracy (z-score) for each COMT genotype across tasks at each of the four loads. There were no significant genotype-by-load effects. Error bars represent ± 1 standard error of the mean.

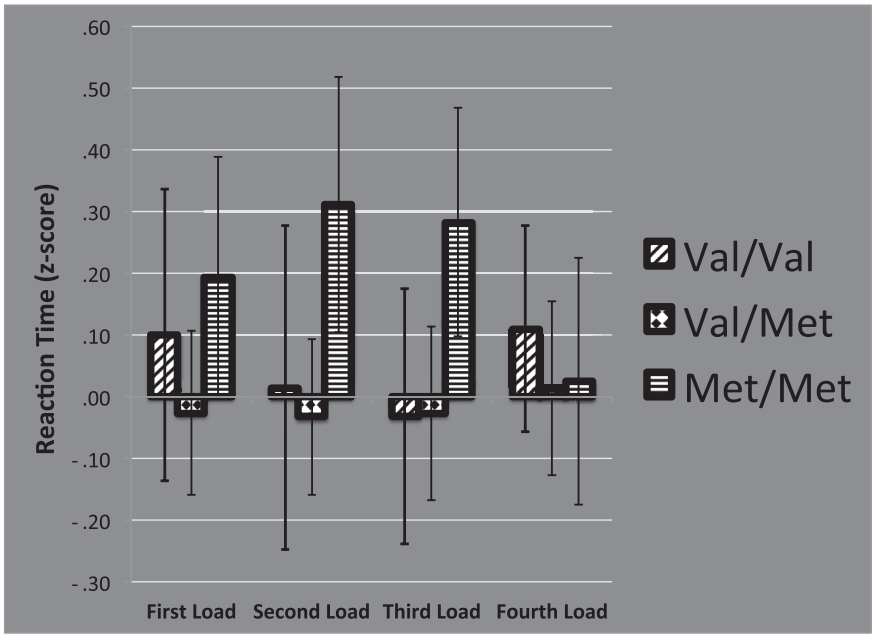

Fig. 5 – Standardized reaction time (z-score) for each COMT genotype across tasks at each of the four loads. There were no significant genotype-by-load effects.

responding) at the 2-back load. This resulted in the exclusion of four participants. One additional participant was removed for extremely low average RT (<100 msec), which was deemed to reflect inappropriate attention to responding, leaving 106 participants in the sample who were retained for the main analyses. There were no significant genotype-by-load effects on accuracy, $F(6, 204) = 1.18$; $p = .316$ (Fig. 6), or reaction time, $F(6, 204) = .53$; $p = .784$ (Fig. 7), on this task, and no main effects of COMT genotype on accuracy, $F(2, 103) = .58$; $p = .562$, or reaction time, $F(2, 103) = 1.20$; $p = .306$. There were expected main effects of load on accuracy, $F(3, 101) = 57.03$; $p < .001$, and reaction time, $F(3, 101) = 5.28$; $p = .002$. Comparing performance between genotypes at each individual load also revealed no significant differences in accuracy (Supplementary Table 3) or reaction time (Supplementary Table 4). Within the high-performing group, which consisted

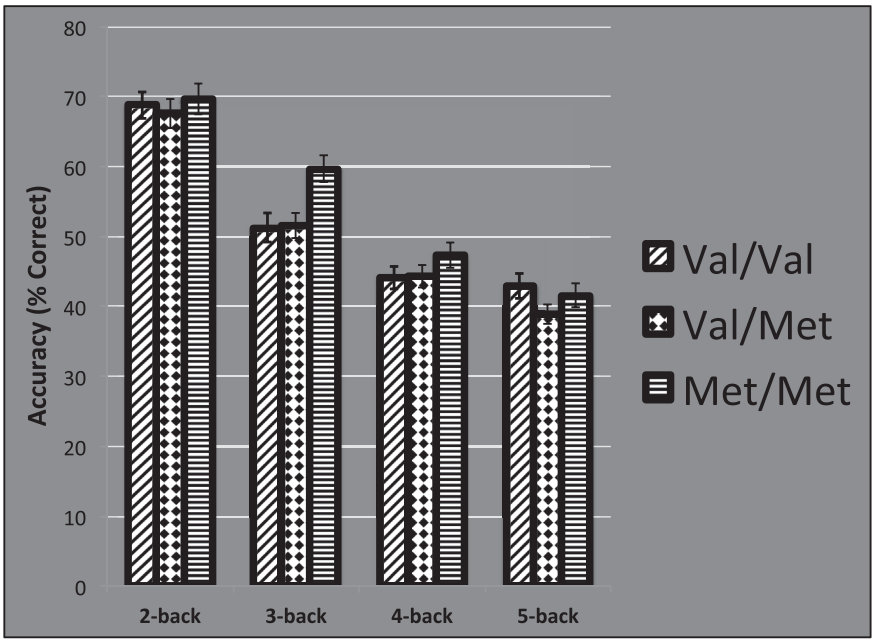

Fig. 6 – Accuracy (percent correct) for each COMT genotype at each of the four loads of the Spatial N-back task. There were no significant genotype-by-load effects.

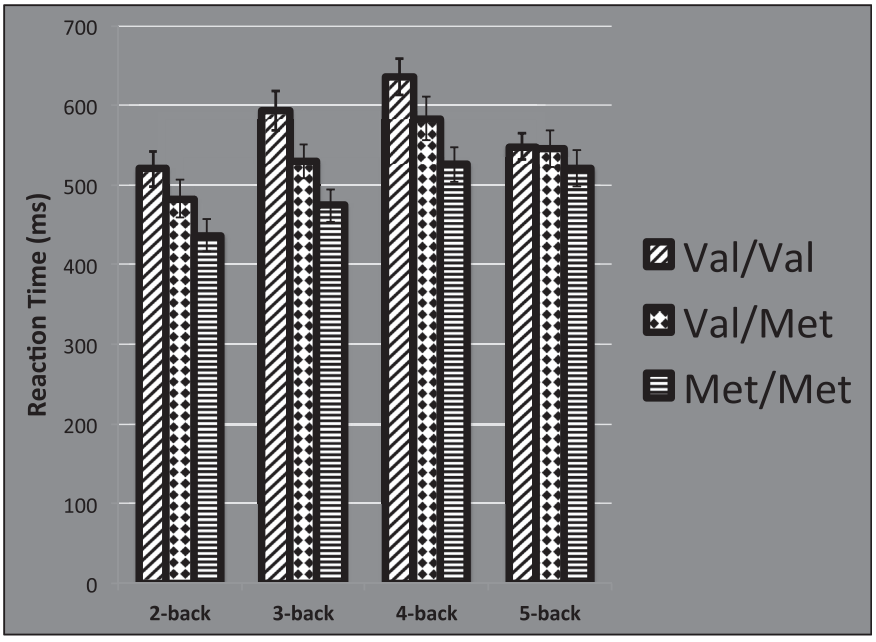

Fig. 7 – Reaction time (milliseconds) for each COMT genotype at each of the four loads of the Spatial N-back task. There were no significant genotype-by-load effects.

of 22 participants who exhibited greater than 90% accuracy for 2-back, there were again no genotype-by-load effects on accuracy $F(6, 57) = .64$; $p = .699$, or reaction time, $F(6, 57) = .83$; $p = .552$. There also were no main effects of COMT on accuracy, $F(2, 19) = .50$; $p = .617$, or reaction time, $F(2, 19) = .10$; $p = .905$. No tests for genotype-by-load interactions or genotype main effects reached significance even at a nominal (uncorrected) threshold of $\alpha = .05$.

4.3. *Sequential N-Back*

Quality control analysis identified two participants who did not perform above the a priori chance rate of accuracy (50% correct responding) on the 2-back load of the Sequential N-back task, and were excluded, leaving 109 participants for the main analyses. There were no significant genotype-by-load effects on accuracy, $F(6, 210) = 1.67$; $p = .131$, or reaction

time, $F(6, 208) = .713$; $p = .640$. Repeated measures ANOVA revealed no main effect of genotype on accuracy, $F(2, 106) = .03$; $p = .967$ (Fig. 8). A nominally significant effect of genotype on reaction time, $F(2, 106) = 4.74$; $p = .011$ (Fig. 9) did not survive correction for multiple comparisons. There were expected main effects of load on accuracy, $F(3, 104) = 16.55$; $p < .001$, and reaction time, $F(3, 103) = 7.00$; $p < .001$. Comparing performance between genotypes at each individual load revealed no significant differences that survived correction for multiple comparisons (Supplementary Tables 5 and 6). In the high-performing group, which consisted of 57 participants who performed above 90% accuracy on the 2-back, there were no significant genotype-by-load effects on accuracy, $F(5.09, 137.45) = 1.39$; $p = .231$, or reaction time, $F(6, 162) = .24$; $p = .965$, and no main effect of genotype on accuracy, $F(2, 54) = .26$; $p = .769$. A nominally significant effect of genotype on reaction time, $F(2, 54) = 3.75$; $p = .030$ did not survive correction for multiple comparisons. The effects of genotype at the uncorrected $\alpha = .05$ level in both the full sample and the high performing group, appeared to indicate *slower* reaction times for Met/Met individuals than individuals in the other genotype groups. Analysis of individual loads indicated that this effect was due to increased response times for Met/Met carriers relative to other genotype groups at 4-back and 5-back. d′ analysis for the sequential N-back, reported in Supplementary Results, yielded a pattern of results similar to the accuracy analysis for this task (Supplementary Figure 2), including no main or interaction effects of genotype.

4.4. *Sternberg task*

Quality control analysis identified six participants who did not perform above the a priori rate of chance accuracy (50% correct responding) on the first or second load of the Sternberg task, and were excluded, leaving 105 participants for the main analyses. The second load was included in quality control criteria because average accuracy remained quite high (>90%). We found no significant genotype-by-load effects on accuracy, $F(6, 202) = .78$; $p = .587$ (Fig. 10), or reaction time, $F(6, 202) = .48$; $p = .867$ (Fig. 11). Again, there were no main effects of COMT genotype on accuracy, $F(2, 102) = .60$; $p = .548$, or reaction time, $F(2, 102) = .54$; $p = .586$. There were expected main effects of load on accuracy, $F(3, 100) = 87.09$; $p < .001$, and reaction time, $F(3, 100) = 26.09$; $p < .001$. When performance was compared between genotypes at each load, there also were no significant differences in accuracy (Supplementary Table 7) or reaction time (Supplementary Table 8). In the high-performing group, which consisted of 96 participants who had above 90% accuracy on the 4-letter load, there were no genotype-by-load effects on accuracy, $F(4.97, 230.86) = .94$; $p = .453$, or reaction time $F(5.02, 233.53) = .71$; $p = .615$, or main effects of genotype on accuracy, $F(2, 93) = .51$; $p = .604$, or reaction time, $F(2, 93) = .05$; $p = .948$. No tests for genotype-by-load interactions or genotype main effects reached significance even at a nominal threshold of $\alpha = .05$.

4.5. *Quantifying and confirming evidence of negative findings for a Met allele advantage in WM*

Given the strong indication of a negative finding with regards to the COMT efficiency hypothesis, we sought to quantify the extent of evidence for the null hypothesis (i.e., the hypothesis that COMT genotype has no effect on WM). Bayes Factors were computed to determine the relative likelihoods of obtaining the observed data given the null hypothesis versus the alternative hypotheses (Jarosz & Wiley, 2014) for between-

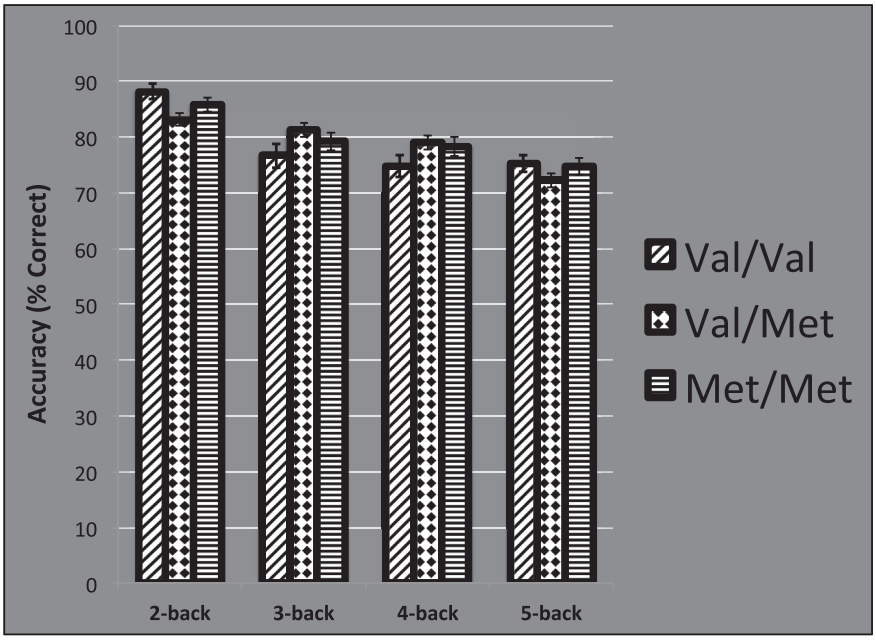

Fig. 8 – Accuracy (percent correct) for each COMT genotype at each of the four loads of the Sequential N-back task. There were no significant genotype-by-load effects.

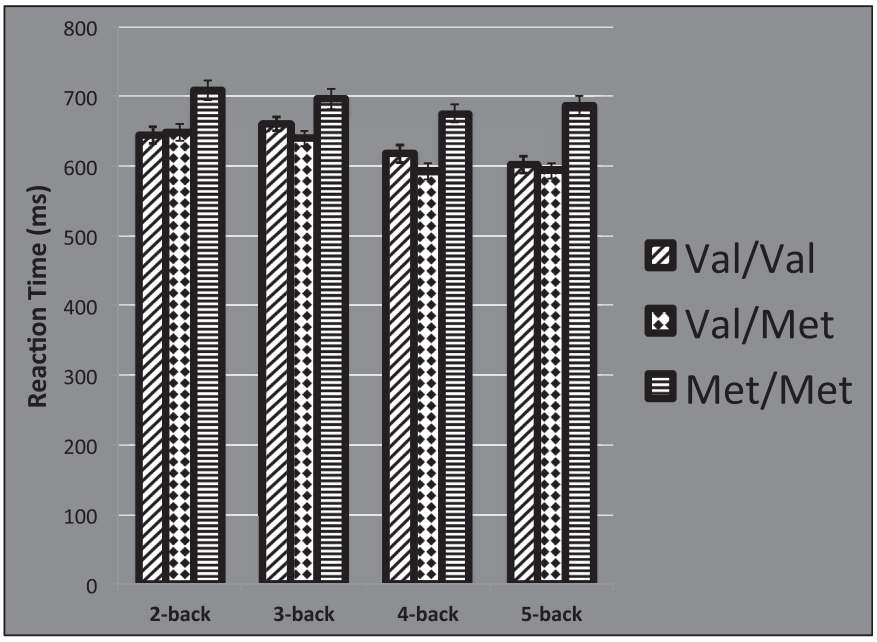

Fig. 9 – Reaction time (msec) for each COMT genotype at each of the four loads of the Sequential N-back task. There were no significant genotype-by-load effects.

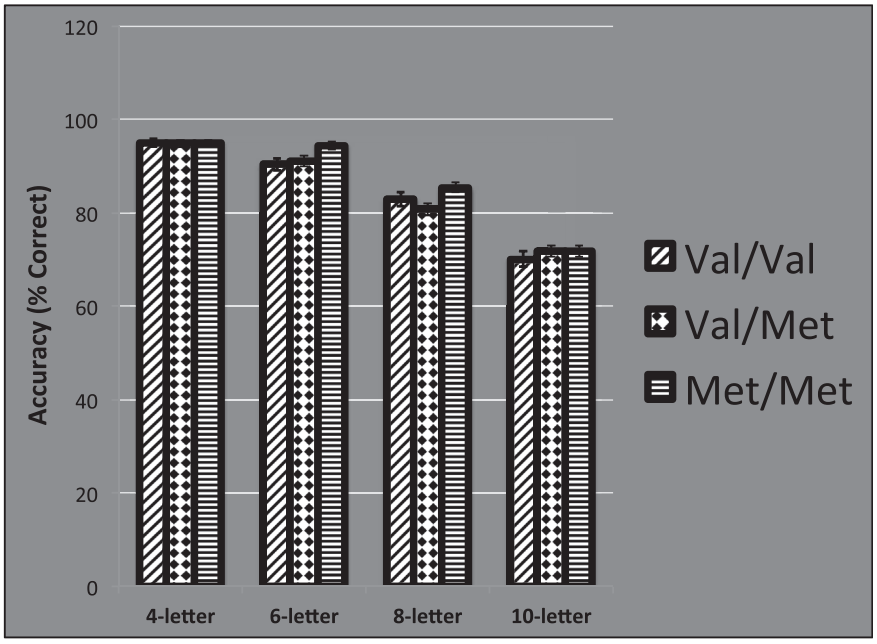

Fig. 10 – Accuracy (percent correct) for each COMT genotype at each of the four loads of the Sternberg task. There were no significant genotype-by-load effects.

genotype models within each task at each load (Supplementary Tables 3–8). For loads on which Met/Met carriers showed nominally highest accuracy, Bayes Factors indicated that the observed data were between 22.7 and 104.6 times more likely to occur given the null hypothesis. For loads on which Met/Met carriers showed nominally lowest RT, Bayes Factors indicated that the observed data were between 20.9 and 93.5 times more likely to occur given the null hypothesis.

Given prior evidence that heterozygous genotype for COMT $Val^{108/158}Met$ yields enzyme activity that is intermediate between Val/Val and Met/Met homozygotes (Chen et al., 2004), we re-coded the COMT genotype variable as ordinal (i.e., 1 for Met/Met, 2 for Val/Met, and 3 for Val/Val) in our ANOVA models, and confirmed that this did not alter the findings. To further explore the possibility that our analysis might be overly conservative, we also ran linear regression models with the ordinal genotype variable to test for any effects of COMT

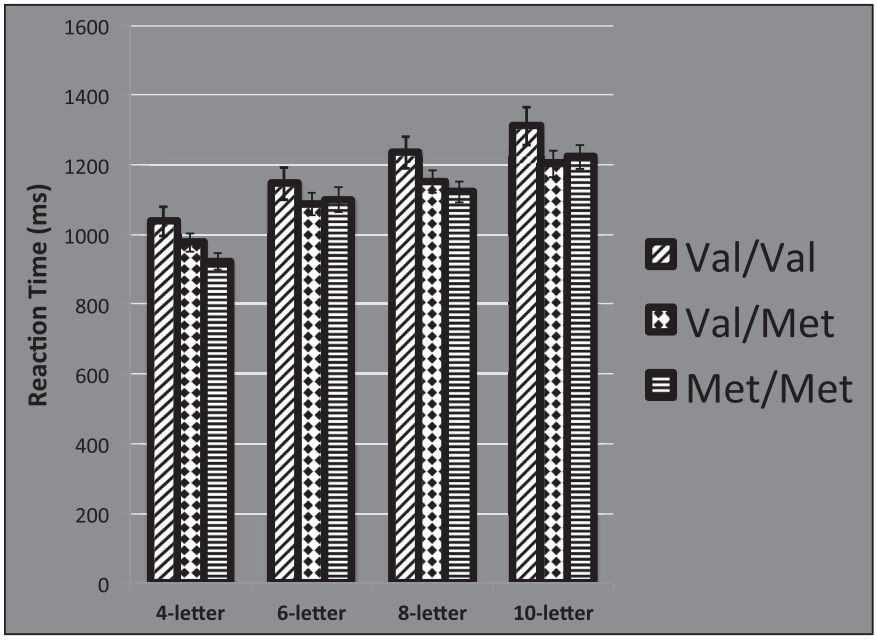

Fig. 11 – Reaction time (msec) for each COMT genotype at each of the four loads of the Sternberg task. There were no significant genotype-by-load effects.

genotypes on average accuracy or reaction time across loads. In regression models for each task and across tasks, there were no significant effects.

5. Discussion

Our goal was to test whether the COMT-Met allele, which is linked to ostensibly more efficient neural activity at low working memory loads (Egan et al., 2001; Mattay et al., 2003; Mier et al., 2010; Tan, Chen, Goldberg, et al., 2007), actually confers a performance advantage when working memory becomes more challenging. If neural efficiency at low loads is associated with more available neural resources for coping with higher demand, increasing demand might reveal latent differences in working memory ability between COMT genotypes. Differences in ability may have been masked by low working memory demand in prior studies that did not find genotype effects on performance. We modified classical working memory tasks to become quite difficult, using loads that increased to higher levels than those tested in previous studies of COMT and working memory. Across three tasks, we found no genotype-by-load interaction effects indicating that, as demands on neural working memory resources increased, this did not lead to increasing effects of COMT genotype on performance (accuracy or reaction time). Notably, there were no main effects of COMT genotype on performance across loads, nor even any effects at any individual load of any individual task that survived multiple comparisons correction – indeed the great majority of tests failed to reach even nominal significance. These findings remained when we analyzed separately the participants who performed with the highest accuracy at the lowest load of each task. The overall lack of behavioral differences between COMT genotypes in the present study is consistent with prior evidence indicating no significant genotype differences on cognitive performance measures at low loads (Dennis et al., 2010; Wardle et al., 2013), and run contrary to a cognitive advantage interpretation of the COMT efficiency account.

An alternative to the cognitive advantage interpretation of COMT efficiency is that differences in activation between COMT genotypes reveal genotype-dependent differences in neural recruitment strategies, which may be comparable in efficacy for meeting working memory demands (if not in neural efficiency), and therefore not directly related to performance, at least in the healthy young brain. Prior evidence has not always indicated that better performance is associated with lower levels of PFC activation. Some prior studies of complex cognitive function that depends on PFC have found better cognitive performance to be associated with greater activation (Gray et al., 2003; Osaka et al., 2004), or greater prefrontal activation in some areas concurrent with less activation in others (Green et al., 2013; Rypma & Prabhakaran, 2009). With regard to COMT genotype, the Met allele has been associated with greater activation in some brain regions and less activation in others, including frontal areas, across multiple studies (Congdon et al., 2009; Winterer et al., 2006; de Frias et al., 2010), and even within the same study (Green et al., 2013). COMT may be a genetic factor impacting neural recruitment more than actual cognitive performance (Dennis et al., 2010; Wardle et al., 2013). While the term 'efficiency' has been used to describe differences in BOLD fMRI measures of brain activation, this phenomenon appears to be underspecified with respect to behavioral significance and underlying mechanisms that result in decreased fMRI BOLD signal (Poldrack, 2014). Satisfactory explanations for differences in activation may need to account for a number of factors, such as whether the overall cognitive processing and precise types

of computations performed in tasks are consistent across all subjects (Poldrack, 2014). Such factors might plausibly impact brain activation without substantially affecting performance outcomes on tasks. Genotype-dependent differences in COMT enzymatic activity appear to contribute to differences in allocation of neural resources during working memory, but this does not necessarily imply substantive differences in working memory ability.

While the present data contribute to an emerging recognition that closely considering the meaning of neural efficiency is important (Poldrack, 2014), the present study focused only on a single question related to behavioral predictions of the COMT efficiency account for working memory. Questions about neural efficiency more broadly are beyond the scope of this study, including, effects of presumed neural efficiency outside the context of COMT genotype, and effects of COMT genotype that extend beyond working memory (e.g., Green et al., 2013). Further investigation, perhaps employing high levels of task difficulty in tasks other than working memory assays, will be informative for developing a broader and more nuanced understanding of these issues.

Perhaps the most compelling aspect of the present findings is the consistency with which COMT genotype effects were not observed across multiple tasks and multiple loads tested. This indication of a null effect is consistent with prior evidence that effects of COMT on behavioral phenotype are small or nonexistent (Barnett et al., 2008; Bruder et al., 2005; Dennis et al., 2010; Flint & Munafo, 2007; Raz et al., 2011; Wardle et al., 2013). This evidence coheres more broadly with an emerging consensus that substantial effects on behavioral phenotype cannot fruitfully be sought via studies of candidate genes, and that the apparent extent of candidate gene evidence has likely been distorted by publication bias (for review) and relevant discussion, see (Flint & Munafo, 2013; Munafo, 2009; Munafo, Clark, & Flint, 2004; Munafo & Flint, 2004, 2009). Evidence in the present data set is bolstered by the calculation of Bayes factors for genotype comparison models. Bayes factors are particularly useful for quantifying the evidence for the null hypothesis, including in candidate gene studies. Rather than representing inferences based on theoretical distributions, they indicate the likelihood of the specific set of observed data given the null or alternative hypothesis (for review) and a practical guide, see (Jarosz & Wiley, 2014). Notably, this approach provides a ratio of likelihoods that can be interpreted similarly across studies of different sample sizes, which is not the case for standard inferences based on *p*-values. This is helpful in the context of behavioral genetic studies not performed with very large datasets. The evidence for the null hypothesis indicated by Bayes factors in the present data thus offers strong support for the absence of an efficiency-related COMT effect on WM.

The three tasks we employed (Spatial N-back, Sequential N-back, and Sternberg Task) measured working memory in different ways. This provided a broader sample of the working memory construct than most prior investigations of COMT effects, and made it particularly notable that none of these tasks showed significant genotype-by-load interaction effects or corrected significant main effects of COMT genotype. In one of the three tasks, the Sequential N-back, nominally slower reaction times were observed in Met/Met individuals. Though it is not possible to draw clear conclusions from a result that did not approach significance after correction for multiple comparisons, it may be worth noting that the observed direction of this nominal genotype effect is not consistent with the Met-associated cognitive advantage interpretation of the COMT efficiency account. The sequential N-back differs in several respects from the spatial N-back, which is the task upon which the COMT efficiency account has been primarily based. The Sequential N-back is not as visually complex and may require less inhibition because subjects only respond to indicate whether or not the present stimulus is the same as the one that appeared n back in the sequence, as opposed to responding to indicate what the stimulus appearing n back actually was. In addition, there is only one type of information available to support correct responding in the Sequential N-back task (i.e., the identity of the item presented n back in the sequence). By contrast, there are two types of information available in the Spatial N-back (i.e., the identity of the item and its location), either or both of which can support successful responding on any given trial. In our sample, and in extensive prior work that we have done with the Spatial N-back task, participants frequently report using both verbal and spatial information. In both of the N-back tasks, there is some indication that participants may have placed greater emphasis on responding quickly on high-loads trials at the expense of accuracy. This could help to explain why reaction time stayed the same or decreased as load increased in the Sequential N-back, and as load increased from 4-back to 5-back in the Spatial N-back. As information load began to challenge the limits of their working memory capacity, participants may have resigned themselves to making their best guess based on less precise memory traces rather than taking time to recall a more precise memory trace with greater certainty. The Sternberg task introduces additional components that differ substantially from either of the two N-back tasks. The Sternberg task requires encoding a large set of stimuli on each trial, rather than a single stimulus, and this set of stimuli must be maintained in working memory during a delay in which, unlike the N-back tasks, no intervening stimuli are presented. Thus, while the Sternberg task clearly taxes the item storage capacity of working memory, the need to inhibit potential distracters during delay is reduced.

These differences between tasks, each of which validly measures the construct of working memory, raise the possibility that some observations of Met-related efficiency may be tied to unique aspects of the Spatial N-back, rather than to working memory more generally. Notably, in a prior study that sought to relate neural efficiency to working memory performance through a repetition training paradigm, the type of information being stored in working memory, in particular spatial versus non-spatial information, impacted the relationship of neural efficiency to performance (Sayala et al., 2006). That study did not investigate genotype, but underscores the importance of considering specific task demands when interpreting the relationship of neural activity to the relatively broad cognitive construct of working memory.

It is possible, though it appears unlikely, that the working memory tasks we employed were not challenging enough at the highest loads to elicit genotype-by-load effects. While average performance was certainly below ceiling levels, and

indeed many participants performed at levels proximate to a priori rates of chance accuracy at the highest loads, average accuracy remained above a priori rates of chance for all task loads. It may be possible to address this issue in future research by increasing working memory load until average performance declines even further than in the present study. It is also possible that the addition of more participants would impact the results, however the general lack of even trend-level effects of COMT genotype on behavioral performance indicates that a larger sample size would be unlikely to yield substantive effects. Follow-up work might also seek to replicate the present paradigm while collecting fMRI data to directly measure levels of neural activity in DLPFC as working memory load increases. Such a study would have to overcome several pragmatic and interpretive hazards related to the required sample size, length and number of tasks, and "noisy" responding as difficulty increases. The present study focused on COMT genotypes with established associations to differential levels of DLPFC activity during working memory in the neuroimaging literature, employed the same measures of working memory used in prior neuroimaging studies, and tested a sample that was demographically similar to prior fMRI cohorts.

Another consideration is that we studied only healthy individuals, and perhaps the consequences of variation in the COMT gene are more related to cognitive function in neuropsychiatric disorders, especially schizophrenia, which involves disordered prefrontal dopaminergic function (Weinberger et al., 2001; Winterer & Weinberger, 2004). The COMT-Val allele has shown slight associations with increased risk for schizophrenia (Egan et al., 2001; Glatt et al., 2003), though other meta-analysis has not supported this conclusion (Munafo et al., 2005) and achieving sufficient statistical power remains a major issue in this literature. There have been some positive reports associating COMT genetic variation with brain activation (Bertolino et al., 2006; Egan et al., 2001) and executive functioning (Bilder et al., 2002; Galderisi et al., 2005) in schizophrenia, but it remains unclear if these were false positives, since none of these findings was reported in large samples or well-replicated. Thus, independent of its relevance to performance measures in healthy individuals, there remains some possibility that COMT-associated differences may be relevant as a neural endophenotype for schizophrenia (Weinberger et al., 2001).

6. Conclusions

While neuroimaging research has shown differences in ostensible prefrontal efficiency between COMT genotypes during working memory, there have been inconsistent findings with regard to associations between COMT and behavioral performance. The present study increased working memory load from relatively easy levels to levels more difficult than those used in prior COMT research. Increasing demand did not yield differences in working memory performance between COMT genotypes. This finding indicates that putative COMT-Met-related prefrontal efficiency may not be directly linked to a behavioral advantage for working memory, even at high levels of difficulty. Rather, variation in COMT genotype may contribute to different, but comparably efficacious, neural recruitment strategies in the healthy young brain. This finding fits within a growing literature indicating little or no effect of COMT on behavioral phenotype, and bolsters the broader argument that substantial effects on phenotype are unlikely to emerge from studies of individual candidate genes.

Acknowledgments

This research was sponsored in part by the National Science Foundation (DRL-1420481 to A. Green), and grants from The American Legacy Foundation, The John Templeton Foundation (ID51971), and Partners in Research at Georgetown University to A. Green, funds from the National Institute of Mental Health Intramural Research program to K. Berman, and graduate student support to J. Ihne from the Georgetown University Office of Biomedical Graduate Education. We thank D. Dickinson for thoughtful comments on data analysis.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2015.10.025.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131*(1), 30–60. http://dx.doi.org/10.1037/0033-2909.131.1.30.

Altamura, M., Elvevåg, B., Blasi, G., Bertolino, A., Callicott, J. H., Weinberger, D. R., et al. (2007). Dissociating the effects of Sternberg working memory demands in prefrontal cortex. *Psychiatry Research Neuroimaging, 154*(2), 103–114.

Apud, J. A., Mattay, V., Chen, J., Kolachana, B. S., Callicott, J. H., Rasetti, R., et al. (2007). Tolcapone improves cognition and cortical information processing in normal human subjects. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 32*(5), 1011–1020.

Axelrod, J., & Tomchick, R. (1958). Enzymatic O-methylation of epinephrine and other catechols. *The Journal of Biological Chemistry, 233*(3), 702–705.

Barch, D. M., Moore, H., Nee, D. E., Manoach, D. S., & Luck, S. J. (2012). CNTRICS imaging biomarkers selection: Working memory. *Schizophrenia Bulletin, 38*(1), 43–52. http://dx.doi.org/10.1093/schbul/sbr160.

Barnett, J. H., Scoriels, L., & Munafò, M. R. (2008). Meta-analysis of the cognitive effects of the catechol-O-methyltransferase gene Val158/108Met polymorphism. *Biological Psychiatry, 64*(2), 137–144.

Bertolino, A., Caforio, G., Petruzzella, V., Latorre, V., Rubino, V., Dimalta, S., et al. (2006). Prefrontal dysfunction in schizophrenia controlling for COMT Val158Met genotype and working memory performance. *Psychiatry Research, 147*(2–3), 221–226.

Bielinski, J., & Davison, M. L. (1998). Gender differences by item difficulty interactions in multiple-choice Mathematics items. *American Education Research Journal, 35*, 455–476.

Bilder, R. M., Volavka, J., Czobor, P., Malhotra, A. K., Kennedy, J. L., Ni, X., et al. (2002). Neurocognitive correlates of the COMT Val(158)Met polymorphism in chronic schizophrenia. *Biological Psychiatry, 52*(7), 701–707.

Bruder, G. E., Keilp, J. G., Xu, H., Shikhman, M., Schori, E., Gorman, J. M., et al. (2005). Catechol-O-methyltransferase (COMT) genotypes and working memory: associations with differing cognitive operations. *Biological Psychiatry, 58*(11), 901–907.

Caldu, X., Vendrell, P., Bartres-Faz, D., Clemente, I., Bargallo, N., Jurado, M. A., et al. (2007). Impact of the COMT Val108/158 Met and DAT genotypes on prefrontal function in healthy subjects. *NeuroImage, 37*(4), 1437–1444.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., et al. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex (New York, N.Y. : 1991), 9*(1), 20–26.

Callicott, J. H., Ramsey, N. F., Tallent, K., Bertolino, A., Knable, M. B., Coppola, R., et al. (1998). Functional magnetic resonance imaging brain mapping in psychiatry: methodological issues illustrated in a study of working memory in schizophrenia. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 18*(3), 186–196.

Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., et al. (2004). Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *American Journal of Human Genetics, 75*(5), 807–821.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*(6625), 604–608.

Congdon, E., Constable, R. T., Lesch, K. P., & Canli, T. (2009). Influence of SLC6A3 and COMT variation on neural activation during response inhibition. *Biological Psychology, 81*(3), 144–152.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*(6554), 279–281.

Dennis, N. A., Need, A. C., LaBar, K. S., Waters-Metenier, S., Cirulli, E. T., Kragel, J., et al. (2010). COMT val108/158 met genotype affects neural but not cognitive processing in healthy individuals. *Cerebral Cortex (New York, N.Y. : 1991), 20*(3), 672–683.

Diaz-Asper, C. M., Goldberg, T. E., Kolachana, B. S., Straub, R. E., Egan, M. F., & Weinberger, D. R. (2008). Genetic variation in catechol-O-methyltransferase: effects on working memory in schizophrenic patients, their siblings, and healthy controls. *Biological Psychiatry, 63*(1), 72–79.

Dingemanse, J., Jorga, K. M., Schmitt, M., Gieschke, R., Fotteler, B., Zürcher, G., et al. (1995). Integrated pharmacokinetics and pharmacodynamics of the novel catechol-O-methyltransferase inhibitor tolcapone during first administration to humans. *Clinical Pharmacology and Therapeutics, 57*(5), 508–517.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience, 28*(16), 4183–4191. http://dx.doi.org/10.1523/JNEUROSCI.0556-08.2008.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., et al. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America, 98*(12), 6917–6922.

Farrell, S. M., Tunbridge, E. M., Braeutigam, S., & Harrison, P. J. (2012). COMT Val(158)Met genotype determines the direction of cognitive effects produced by catechol-O-methyltransferase inhibition. *Biological Psychiatry, 71*(6), 538–544.

Flint, J., & Munafo, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine, 37*(2), 163–180. http://dx.doi.org/10.1017/S0033291706008750.

Flint, J., & Munafo, M. R. (2013). Candidate and non-candidate genes in behavior genetics. *Current Opinion in Neurobiology, 23*(1), 57–61. http://dx.doi.org/10.1016/j.conb.2012.07.005.

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology General, 144*, 30–47.

de Frias, C. M., Marklund, P., Eriksson, E., Larsson, A., Oman, L., Annerbrink, K., et al. (2010). Influence of COMT gene polymorphism on fMRI-assessed sustained and transient activity during a working memory task. *Journal of Cognitive Neuroscience, 22*(7), 1614–1622.

Galderisi, S., Maj, M., Kirkpatrick, B., Piccardi, P., Mucci, A., Invernizzi, G., et al. (2005). Catechol-O-methyltransferase Val158Met polymorphism in schizophrenia: associations with cognitive and motor impairment. *Neuropsychobiology, 52*(2), 83–89.

Giakoumaki, S. G., Roussos, P., & Bitsios, P. (2008). Improvement of prepulse inhibition and executive function by the COMT inhibitor tolcapone depends on COMT Val158Met polymorphism. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 33*(13), 3058–3068.

Gladwin, T. E., den Uyl, T. E., Fregni, F. F., & Wiers, R. W. (2012). Enhancement of selective attention by tDCS: interaction with interference in a Sternberg task. *Neuroscience Letters, 512*(1), 33–37. http://dx.doi.org/10.1016/j.neulet.2012.01.056.

Glatt, S. J., Faraone, S. V., & Tsuang, M. T. (2003). Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: meta-analysis of case-control and family-based studies. *American Journal of Psychiatry, 160*(3), 469–476.

Goldman-Rakic, P. S., Muly, E. C., & Williams, G. V. (2000). D(1) receptors in prefrontal cells and circuits. *Brain Research. Brain Research Reviews, 31*(2–3), 295–301.

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage, 38*(2), 346–356. http://dx.doi.org/10.1016/j.neuroimage.2007.07.041.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*(3), 316–322.

Green, A. E., Kraemer, D. J., DeYoung, C. G., Fossella, J. A., & Gray, J. R. (2013). A gene-brain-cognition pathway: prefrontal activity mediates the effect of COMT on cognitive control and IQ. *Cerebral Cortex (New York, N.Y. : 1991), 23*(3), 552–559.

Green, A. E., Munafo, M. R., DeYoung, C. G., Fossella, J. A., Fan, J., & Gray, J. R. (2008). Using genetic data in cognitive neuroscience: from growing pains to genuine insights. *Nature Reviews Neuroscience, 9*, 710–720.

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cerebral Cortex, 13*(12), 1369–1374.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? a practical guide to computing and reporting Bayes factors. *Journal of Problem Solving, 7*, 2–9.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*(4), 462–475.

Käenmäki, M., Tammimäki, A., Myöhänen, T., Pakarinen, K., Amberg, C., Karayiorgou, M., et al. (2010). Quantitative role

of COMT in dopamine clearance in the prefrontal cortex of freely moving mice. *Journal of Neurochemistry, 114*(6), 1745–1755.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology General, 132*(1), 47–70.
Lachman, H. M., Papolos, D. F., Saito, T., Yu, Y. M., Szumlanski, C. L., & Weinshilboum, R. M. (1996). Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics, 6*(3), 243–250.
Lee, L. O., & Prescott, C. A. (2014). Association of the catechol-O-methyltransferase val158met polymorphism and anxiety-related traits: a meta-analysis. *Psychiatric Genetics, 24*(2), 52–69.
Linden, D. E., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., et al. (2003). Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. *NeuroImage, 20*(3), 1518–1530.
Lotta, T., Vidgren, J., Tilgmann, C., Ulmanen, I., Melén, K., Julkunen, I., et al. (1995). Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry, 34*(13), 4202–4210.
Mattay, V. S., Goldberg, T. E., Fera, F., Hariri, A. R., Tessitore, A., Egan, M. F., et al. (2003). Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences of the United States of America, 100*(10), 6186–6191.
McNicol, D. (1972). *A primer of signal detection theory*. London: Allen & Unwin.
Mier, D., Kirsch, P., & Meyer-Lindenberg, A. (2010). Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Molecular Psychiatry, 15*(9), 918–927.
Munafo, M. R. (2009). Reliability and replicability of genetic association studies. *Addiction, 104*(9), 1439–1440. http://dx.doi.org/10.1111/j.1360-0443.2009.02662.x.
Munafo, M. R., Bowes, L., Clark, T. G., & Flint, J. (2005). Lack of association of the COMT (Val158/108 Met) gene and schizophrenia: a meta-analysis of case-control studies. *Molecular Psychiatry, 10*(8), 765–770.
Munafo, M. R., Clark, T. G., & Flint, J. (2004). Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. *Psychiatry Research, 129*(1), 39–44. http://dx.doi.org/10.1016/j.psychres.2004.06.011.
Munafo, M. R., & Flint, J. (2004). Meta-analysis of genetic association studies. *Trends in Genetics, 20*(9), 439–444. http://dx.doi.org/10.1016/j.tig.2004.06.014.
Munafo, M. R., & Flint, J. (2009). Replication and heterogeneity in gene x environment interaction studies. *International Journal of Neuropsychopharmacology, 12*(6), 727–729. http://dx.doi.org/10.1017/S1461145709000479.
Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage, 11*(5 Pt 1), 424–446.
Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *NeuroImage, 21*(2), 623–631.
Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46–59.
Palmatier, M. A., Kang, A. M., & Kidd, K. K. (1999). Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. *Biological Psychiatry, 46*(4), 557–567.
Poldrack, R. A. (2015). Is "efficiency" a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience, 11*, 12–17.
Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., et al. (2001). Gating of human theta oscillations by a working memory task. *Journal of Neuroscience, 21*(9), 3175–3183.
Raz, N., Dahle, C. L., Rodrigue, K. M., Kennedy, K. M., & Land, S. (2011). Effects of age, genes, and pulse pressure on executive functions in healthy adults. *Neurobiology of Aging, 32*(6), 1124–1137.
Rypma, B., & Prabhakaran, V. (2009). When less is more and when more is more: the mediating roles of capacity and speed in brain-behavior efficiency. *Intelligence, 37*(2), 207–222.
Sayala, S., Sala, J. B., & Courtney, S. M. (2006). Increased neural efficiency with repeated performance of a working memory task is information-type dependent. *Cerebral Cortex, 16*(5), 609–617.
Sesack, S. R., Hawrylak, V. A., Matus, C., Guido, M. A., & Levey, A. I. (1998). Dopamine axon varicosities in the prelimbic division of the rat prefrontal cortex exhibit sparse immunoreactivity for the dopamine transporter. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 18*(7), 2697–2708.
Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.), 153*(3736), 652–654.
Tan, H., Chen, Q., Goldberg, T. E., Mattay, V. S., Meyer-Lindenberg, A., Weinberger, D. R., et al. (2007). Catechol-O-methyltransferase Val158Met modulation of prefrontal-parietal-striatal brain systems during arithmetic and temporal transformations in working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 27*(49), 13393–13401.
Tan, H., Chen, Q., Sust, S., Buckholtz, J. W., Meyers, J. D., Egan, M. F., et al. (2007). Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes on working memory brain function. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12536–12541.
Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751–754. http://dx.doi.org/10.1038/nature02466.
Veltman, D. J., Rombouts, S. A., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *NeuroImage, 18*(2), 247–256.
Wardle, M. C., de Wit, H., Penton-Voak, I., Lewis, G., & Munafò, M. R. (2013). Lack of association between COMT and working memory in a population-based cohort of healthy young adults. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 38*(7), 1253–1263.
Weinberger, D. R., Egan, M. F., Bertolino, A., Callicott, J. H., Mattay, V. S., Lipska, B. K., et al. (2001). Prefrontal neurons and the genetics of schizophrenia. *Biological Psychiatry, 50*(11), 825–844.
Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Front Psychol, 4*, 433.
Williams, G. V., & Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature, 376*(6541), 572–575.
Winterer, G., Musso, F., Vucurevic, G., Stoeter, P., Konrad, A., Seker, B., et al. (2006). COMT genotype predicts BOLD signal

and noise characteristics in prefrontal circuits. *NeuroImage, 32*(4), 1722–1732.

Winterer, G., & Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences, 27*(11), 683–690.

Zakrzewska, M. Z., & Brzezicka, A. (2014). Working memory capacity as a moderator of load-related frontal midline theta variability in Sternberg task. *Frontiers in Human Neuroscience, 8*, 399. http://dx.doi.org/10.3389/fnhum.2014.00399.